Supplemental Sections

Supplemental Section S1

Isolation of icSARS-CoV and icSARS-CoV ΔORF6. SARS-CoV is unique among human coronaviruses not only for high virulence and severe pathogenic outcomes but also for the large number of accessory open reading frames (ORF) encoded in its genome. Using reverse genetics, these accessory ORF were primary targets for deletion to determine potential roles in SARS-CoV pathogenesis (**Figure S1**). We and others have previously demonstrated that SARS-CoV ORF6: 1) is an interferon antagonist (8), 2) binds to karyopherin α2 and karyopherin α2/karyopherin β1 complexes, sequestering them onto intracellular membranes (**Figure 1A**) (4), 3) prevents STAT1 complex nuclear translocation, an essential step involved in interferon signaling (4) and 4) enhances early viral RNA synthesis (23). To study the role of the ORF6 protein on host expression networks (result of modeling approaches to determine the connectivity/relatedness of genes detected by microarray analysis example of network schematic in Figure 3), we designed a recombinant virus that lacked ORF6, icSARS-CoV ΔORF6 (**Figure S1**). For each virus, plasmids containing the appropriate genome fragments were amplified in E. coli, isolated by restriction digestion, ligated and full length genomic RNA transcribed and electroporated into permissive cells. Viruses were successfully rescued from both the wild type (icSARS-CoV infectious clone derived) and icSARS-CoV ΔORF6 (infectious clone derived virus identical to wild type with the exception of the deletion of ORF6) and each was plaque purified and sequence confirmed prior to use in subsequent experiments. Growth kinetics analysis (MOI of 1)

of icSARS-CoV and icSARS-CoV ΔORF6 in several human and non-human primate cells revealed no significant differences in titer at any time point examined (22). Subsequent studies using another recombinant SARS-CoV with an ORF6 deletion demonstrated that at a low MOI (MOI 0.01) in non-human primate kidney cells that 1 to 5 fold reductions in viral titers could be detected between 6 and 24 hours post infection, but not at later times post infection (23). icSARS-CoV and icSARS-CoV ΔORF6 infected human lung cells were then probed to determine the effects on global transcriptional following the removal of ORF6.

Supplemental Section S2: Proteomic Methods

Proteomics reagents and sample preparation and strong cation exchange fractionation. All chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless stated otherwise. Ammonium bicarbonate and acetonitrile were purchased from Fisher Scientific (Pittsburgh, PA), and sequencing-grade modified trypsin was purchased from Promega (Madison, WI). Bicinchoninic acid (BCA) assay reagents and standards were obtained from Pierce (Rockford, IL); purified, deionized water, >18 MΩ, (Nanopure Infinity ultrapure water system, Barnstead, Dubuque, IA) was used to make all aqueous buffers.

Lysate protein concentrations were determined by BCA protein assay and diluted to a uniform final volume in 50 mM ammonium bicarbonate pH 7.8. Proteins were reduced with 10 mM dithiothreitol, followed by alkylation of free sulfhydryl groups with 40 mM iodoacetamide in the dark; each reaction was performed for 1 h at 37°C with constant shaking at 800 rpm. Denatured and reduced samples were diluted 10-fold with

50 mM ammonium bicarbonate pH 7.8, and $CaCl_2$ was added to a final concentration of 1 mM prior to enzymatic digestion. Sequencing-grade modified trypsin was activated by adding 20 µL of 50 mM ammonium bicarbonate pH 7.8 to 20 µg lyophilized trypsin and incubating for 10 min at 37˚C. Activated trypsin was then added to the samples at 1:50 (w/w) trypsin-to-protein ratio, and samples were digested at 37˚C for 3 h with constant shaking at 800 rpm; reactions were quenched by rapid freezing in liquid nitrogen. Digested samples were desalted using solid phase extraction columns (Discovery C18, Supelco, Bellefonte, PA) according to the manufacturer's instructions. Samples were then concentrated to 100 µL *in vacuo* (Speed-Vac SC 250 Express, Thermo Savant, Holbrook, NY), and a BCA protein assay was performed to verify final peptide concentrations. Samples were stored at –80˚C until strong cation exchange fractionation with liquid chromatography-tandem mass spectrometry (LC-MS/MS) or quantitative LC-MS analyses. Strong cation exchange fractionation was performed as previously described (12, 16). Twenty-four fractions were collected from minute 30 to minute 65 of the gradient, and they were subsequently dried *in vacuo* and stored at -80˚C until LC-MS/MS analysis.

Reversed-phase capillary LC-MS/MS and LC-MS analyses. LC-MS/MS analysis was used to generate an accurate mass and time (AMT) tag database for virus-infected Calu3 2B4 cells (see below). For this, dried peptide fractions were reconstituted in 30 µL of 25 mM ammonium bicarbonate pH 7.8 and analyzed using a 4-column custom-built capillary LC system coupled online to a linear ion trap mass spectrometer (LTQ; Thermo Scientific, San Jose, CA) by way of an in-house manufactured electrospray

ionization interface, as previously described (10). To identify the eluting peptides, the LTQ was operated in a data-dependent MS/MS mode as previously described (12).

Following Calu3 2B4/virus AMT tag database generation, LC-MS analyses were performed on all icSARS-CoV, icSARS-CoV ΔORF6, and mock-infected samples to generate quantitative data. For this, dried peptide samples were reconstituted in 30 µL of 25 mM ammonium bicarbonate, pH 7.8 and analyzed in triplicate and random order using identical chromatographic and electrospray conditions as for LC-MS/MS analyses. The LC system was interfaced to an Exactive mass spectrometer (Thermo Scientific), and the temperature of the heated capillary and the ESI voltage were 250°C and 2.2 kV, respectively. Data were collected over the mass range 400-2,000 *m/z*.

Development of the AMT tag database for virus-infected Calu-3 cells. A novel AMT database was generated for Calu-3 cells using mock-infected and virus infected cells. To generate the AMT tag database, aliquots of the icSARS-CoV, icSARS-CoV ΔORF6 or mock-infected samples were combined to make the following pools: 1) mock-infection (all time points), 2) early SARS-CoV infection (0, 12, 24, and 36 h), and 3) late SARS-CoV infection (48, 54, 60, and 72 h). Each pool was subjected to fractionation, and each fraction was analyzed by LC-MS/MS. The SEQUEST analysis software (3) was used to match the MS/MS fragmentation spectra with sequences from the April 4, 2010 UniProt/Swiss-Prot protein database, containing 20,276 entries (protein TITIN_Human was removed due to excessive length). When searching, SEQUEST used a dynamic mass modification on methionine residues corresponding to oxidation (15.9949 Da) and a static mass modification on cysteinyl residues to account for alkylation by

iodoacetamide (57.0215 Da). Peptides passing the following filter criteria were stored as AMT tags in a Microsoft SQL Server database: 1) SEQUEST DelCn2 value (normalized Xcorr difference between the top scoring peptide and the second highest scoring peptide in each MS/MS spectrum) ≥ 0.10 and 2) SEQUEST correlation score (Xcorr) ≥ 2, 2.6, and 3.5 for fully tryptic peptides with 1+, 2+, and 3+ charge states, respectively, and Xcorr ≥ 2.5, 3.6, and 4.1 for partially tryptic or non-tryptic protein terminal peptides with 1+, 2+, and 3+ charge states, respectively. Non-tryptic peptides were excluded, and a minimum peptide length of 6 amino acid residues was required. These criteria resulted in 56,220 peptides identified with an estimated false discovery rate <2% based on a target-decoy database search (18). The elution times for these peptides were normalized to a range of 0 to 1 using a predictive peptide LC normalized elution time (NET) model and linear regression, as previously reported (15). A NET average and standard deviation were assigned to each identified peptide if the same peptide was observed in multiple analyses. Both calculated monoisotopic masses and observed NETs of identified peptides were included in the AMT tag database.

Processing of quantitative LC-MS datasets. Quantitative LC-MS datasets were processed using the PRISM Data Analysis system (7), which is a series of software tools developed in-house (e.g. Decon2LS (6) and VIPER (13) freely available at http://ncrr.pnl.gov/software/). Individual steps in this data processing approach are reviewed here (24). The peptide identities of detected features in each dataset (here a dataset is equivalent to a single LC-MS analysis) were determined by comparing their measured monoisotopic masses and NETs to the calculated monoisotopic masses and

observed NETs of each of the 56,220 peptides in the filtered AMT tag database within initial search tolerances of ± 6 ppm and ± 0.025 NET for monoisotopic mass and elution time, respectively. The peptides identified from this matching process were retained as a matrix for subsequent data analysis.

Proteomic data processing. The peak intensity values (i.e. abundances) for the final peptide identifications were processed in a series of steps using MatLab® R2010b. Briefly, quality control processing was performed to remove peptides with insufficient data (21) or LC-MS runs showing significant deviation from standard behavior (11). Peptides were normalized across the technical replicates using a rank invariant subset of peptides (p-value threshold of 0.1) followed by mean centering of the data based on the SPANS procedure (20). Proteins were quantified using a standard R-Rollup method (17) using the most abundant reference peptide, after filtering the peptides that were redundant, had low data content, or were outside the dominant significance pattern. Comparative statistical analyses of time-matched mock with icSARS-Cov and icSARS-CoV ΔORF 6 samples was performed using a Dunnett adjusted t-test to assess differences in protein average abundance, and a G-test to assess associations among factors due to the presence/absence of response.

Supplemental Section S3: Modeling Workflow

Goal: To use a system's biology approach to investigate whether the ORF6 accessory protein mediates a specific or more global block of karyopherin-mediated nuclear translocation and host gene expression.

Approach: To evaluate whether transcription factors whose transport is regulated by karyopherins are enriched by the SARS mutation ΔORF6 (i.e. over-connected to downstream genes in the dataset) more than would be expected by chance, we have identified karyopherin-mediated transcriptional hubs in Calu-3 cells based on global transcriptomics from wild-type icSARS-CoV (WT) and mutant icSARS-CoV ΔORF6. Further, we have independently confirmed these targets of karyopherin-mediated nuclear transport in two additional datasets, including global proteomics from Calu-3 cells and global transcriptomics from HAE cells. The steps outlined below are summarized by the schematic in **Supplemental Figure S2**.

Calu-3 genes:

1. **Calu-3 cell global transcriptomics in WT and ΔORF6.** To determine the pattern of differential gene expression for icSARS-CoV and icSARS-CoV ΔORF6 infected and mock-infected cells, Calu3 2B4 cells were plated in triplicate and collected at 0, 3, 7, 12, 24, 30, 36, 48, 54, 60, and 72 hours post infection. RNA isolation from Calu3 2B4 cells and the subsequent Agilent microarray processing was performed as described previously (9).

2. **Filter for genes differentially expressed between WT and ΔORF6.** Differential expression was determined by comparing icSARS-CoV and icSARS-CoV ΔORF6 infected replicates to mock-infected replicates for each time point, based on a linear model fit for each transcript. Criteria for differential expression were an absolute log2 fold change of 1.5 and a False Discovery Rate (FDR) adjusted p-value < 0.05 for a given time point. Differential expression was also calculated directly between icSARS-

CoV and icSARS-CoV ΔORF6 for each time point using criteria of 2-fold change and an FDR adjusted p-value <0.05. Significant transcript values were transformed for clustering and network analysis to fold change (log2) of icSARS-CoV or icSARS-CoV ΔORF6 infected compared to time-matched mock-infected samples.

3. Hierarchical clustering (Figure 2). To determine the significant biological processes associated with icSARS-CoV ΔORF6 differential gene expression, the dataset was first reduced to six clusters by K-means based on common patterns of expression across genes (**Figure 2A**).

4. Identification of gene cluster up in ΔORF6 and not regulated in WT (Figure 2, C4). The pattern of expression in gene cluster C4, which included 1,674 genes upregulated in icSARS-CoV ΔORF6 but unchanged in icSARS-CoV and enriched in biological processes for chromosome organization and regulation of gene expression and nucleosome assembly, was particularly interesting given one of the ORF6 mechanisms of action (**Figures 1A and 2A**). Therefore, we focused on the genes in cluster C4 to identify potential downstream targets of karyopherins whose transcription was blocked in icSARS-CoV infection.

5. Identification of significant biological processes describing difference between WT and ΔORF6 (Figure 2A). To determine the significant biological processes associated with icSARS-CoV ΔORF6 differential gene expression, the dataset was first reduced to six clusters by K-means based on common patterns of expression across genes (**Figure 2A**) and then significant enrichment ($p<0.05$) of biological process Gene Ontology categories was calculated for each cluster individually

(**Figure 2A**). Functional enrichment statistics were determined using DAVID (2, 5) to identify the most significant processes affected by infection. The DAVID functional annotation tool utilizes the Fisher Exact test to measure gene enrichment in biological process Gene Ontology (GO) category terms for significant genes compared to background, which included all genes on the Agilent platform that passed QC criteria.

6. Transcription factor enrichment analysis for karyopherin targets (Tables 1 and 2). To identify major transcriptional regulators whose nuclear import is controlled by karyopherins, the statistical Interactome tool was used in MetaCore to measure the interconnectedness of genes in the experimental dataset relative to all known interactions in the background dataset. Statistical significance of over-connected interactions was calculated using a hypergeometric distribution, where the p value represents the probability of a particular mapping arising by chance for experimental data compared to the background (14). From a biological perspective, the enrichment scores (or p values) represent how connected the transcription factors are to downstream targets (genes or proteins) in the dataset compared to the number of targets expected by chance. If a transcriptional hub is significantly ($p<0.05$) more connected to the dataset, then it suggests that signaling through this hub is uniquely regulated based on the experimental conditions, in this case through mutation of ORF6. In order to determine the consequence of removal of the nuclear import block in SARS-ΔORF6 infection, significantly over-connected transcription factors were filtered for those whose transport is regulated by karyopherins in the cell. When the number of connections between target genes in the dataset and upstream regulatory transcription factors are greater than would be expected by chance (based on the number of known

207 connections), this indicates that these transcription factors are enriched in our dataset
208 and are important regulators of the host response by ORF6. All of the transcription
209 factor hubs (schematically represented by large circles in **Figure 3**) associated with
210 gene cluster C4 that were identified as being regulated by karyopherins in MetaCore
211 (the type of karyopherin is indicated by column heading) are listed in **Table 1**. Six of
212 these transcription factors (vitamin D receptor-VDR, cyclic AMP receptor binding protein
213 1-CREB1, Oct3/4, hypoxia inducible factor α2-Epas1/HIFα2, p53, and SMAD4) were
214 significantly ($p<0.05$) overconnected to the dataset (**Table 1**) indicating that there was
215 differential signaling through these transcription factor hubs when comparing icSARS-
216 CoV ΔORF6 to icSARS-CoV.

217 The target analysis above, which identified the six transcription factors described in
218 **Figures 3 and 4,** was restricted to the subset of genes in Cluster C4 (**Figure 2A**) that
219 were upregulated exclusively in icSARS-ΔORF6 infected cells and was based on the
220 likely scenario that the presence of ORF6 in wild type infected cells blocks transcription
221 of genes that are upregulated in icSARS-ΔORF6 as there is no ORF6 to sequester the
222 karyopherins. However, we wondered if additional karyopherin regulated transcription
223 factors were also present in networks (result of modeling approaches to determine the
224 connectivity/relatedness of genes detected by microarray analysis example in Figure 3)
225 generated from the 6,947 genes (in all 6 clusters) originally identified (**Figure 2A**), which
226 would suggest that karyopherins import transcription factors that both positively and
227 negatively regulate downstream target genes during the course of SARS-CoV infection.
228 To answer this question, we expanded our transcription factor analysis to include all
229 significant genes in the dataset, resulting in 27 over-connected transcription factors

(transcription factors with a significant number of downstream target genes that were differentially expressed, in this case either up or down regulated) whose nuclear importation was mediated by karyopherins (**Table 2**). This analysis provided a more global understanding of the potential impact of the nuclear importation block during SARS-CoV infection compared to the targeted approach described in **Figures 3 and 4**. For example, STAT1 was also identified as an important karyopherin-mediated regulator of gene expression during infection.

7. Network analysis of karyopherin-mediated transcriptional hubs (Figures 3 and 4). Networks were constructed in MetaCore for experimental data using an algorithm that identifies the shortest path to directly connect nodes in the dataset to transcription factors. Network visualizations were created in MetaCore or Cytoscape (19). Over 350 gene nodes in the C4 cluster directly interact with these six transcription factors (**Figure 3**) based on information in the Metacore knowledgebase suggesting their transcription is specifically blocked in the presence of functional ORF6. The network in **Figure 3** was built by identifying the connectivity of genes identified by microarray analysis and then determining the transcription factors responsible for their expression. For example, VDR and CREB1 are both upregulated genes identified as part of the interconnected genes in Cluster C4 and are transcription factors (with many of their downstream regulated genes in the network) that have karyopherin mediated nuclear transport, adding more weight to our modeling data. Some of the genes in the VDR and CREB1 networks whose pattern of expression were confirmed by HAE transcriptomic and Calu3 2B4 proteomic data (**Figures 5 and 6**, respectively) are shown in the enlarged insert of **Figure 3**.

Calu-3 proteins:

8. Calu-3 cell global proteomics in WT and ΔORF6. To determine the pattern of differential protein expression for icSARS-CoV and icSARS-CoV ΔORF6 infected and mock-infected cells, Calu3 2B4 cells were plated in triplicate and collected at 0, 3, 7, 12, 24, 30, 36, 48, 54, 60, and 72 hours post infection. Cell lysates were analyzed by LC-MS/MS as described in **Supplemental Section S1**.

9. Filter for proteins differentially expressed between WT and ΔORF6. Comparative statistical analyses of time-matched mock with icSARS-Cov and icSARS-CoV ΔORF 6 samples was performed using a Dunnett adjusted t-test to assess differences in protein average abundance, and a G-test to assess associations among factors due to the presence/absence of response. A total of 864 proteins, which were significantly ($p<0.05$) different between icSARS-CoV and icSARS-CoV ΔORF6 infected Calu3 2B4 across all time points by either test, were used for transcription factor enrichment analysis.

10. Transcription factor enrichment analysis for karyopherin targets (Table 2). To identify major transcriptional regulators whose nuclear import is controlled by karyopherins, transcription factor enrichment analysis was performed on Calu-3 proteomics as described in step #6. To most directly determine how the proteomics contributes to our understanding of the role for ORF6 in karyopherin-mediated nuclear translocation and host gene expression, we first integrated the transcriptomic and proteomic datasets. From the integrated data, we have a more comprehensive view of the changes mediated by ΔORF6 in Calu-3 cells at both the gene and protein level that allow us to answer two important questions. First, we can determine whether

276 karyopherin-mediated transcriptional hubs are further enriched (i.e. more significant) in
277 the presence of the proteomic data. For example, if the proteomic data supports a role
278 for ORF6 in karyopherin-mediated nuclear transport, then we would expect an increase
279 in the enrichment scores of these hubs after addition of the proteomics; otherwise, the
280 values would decrease. From a biological perspective, the enrichment scores (or p
281 values) represent how connected the transcription factors are to downstream targets
282 (genes or proteins) in the dataset compared to the number of targets expected by
283 chance. If a transcriptional hub is significantly ($p<0.05$) more connected to the dataset,
284 then it suggests that signaling through this hub is uniquely regulated based on the
285 experimental conditions, in this case through mutation of ORF6. Second, we can use
286 the integrated gene and protein data to also identify novel hubs whose downstream
287 targets may only be detected at the protein level. In our study, transcription factor
288 analysis of the combined Calu3 2B4 proteomics and transcriptomics data resulted in an
289 increase in enrichment scores, ranging from 1.5 to $2x10^{12}$-fold, for 22 out of 28 of the
290 transcription factor hubs requiring karyopherin for nuclear importation, including 5 out of
291 the 6 transcription factor hubs uniquely upregulated during icSARS-CoV ΔORF6
292 infection (**Table 2**). This suggests that many protein nodes in the proteomic dataset
293 share common upstream transcription factor regulators with gene nodes in the
294 transcriptional data, including those that require karyopherin for nuclear import. Four
295 transcription factor hubs in particular, C-myc, Rela, specificity protein 1 (Sp1) and
296 STAT1, had p-values from the combined transcriptomic/proteomic analysis that were
297 >1,000-fold more significant than for the transcriptomics alone demonstrating the added
298 value of including targets from both data types (marked with ** in **Table 2**).

HAE genes:

11. HAE cell global transcriptomics in WT and ΔORF6. To determine the pattern of differential gene expression for icSARS-CoV and icSARS-CoV ΔORF6 infected and mock-infected cells, HAE cells were plated in triplicate and collected at 2, 24, 48, and 72 hours post infection. RNA isolation from HAE cells was performed as described previously (9). Equivalent amounts of RNA from three biological replicates from each condition were pooled. Microarray analysis was performed as previously described (1) using Agilent 4x44K Whole Human Gene Expression Microarrays.

12. Filter for genes differentially expressed between WT and ΔORF6. Genes were filtered for differential expression between icSARS-CoV and icSARS-CoV ΔORF6 infected cells based on a 2-fold change cut-off.

13. Transcription factor enrichment analysis for karyopherin targets (Results; Validation of transcription factor hubs). To identify major transcriptional regulators whose nuclear import is controlled by karyopherins, transcription factor enrichment analysis was performed on HAE transcriptomics as described in step #6. Transcription factor analysis of the HAE dataset resulted in 7 enriched transcription factor hubs regulated by karyopherins ($p<0.05$), all of which overlap with the Calu3 2B4 dataset. These transcription factor hubs included RelA, C-jun, CREB1, Hif1α, C-fos, VDR and SMAD3.

Validation:

14. Comparison of karyopherin-mediated transcriptional hubs across datasets (Table 2 and Figure 5). To independently confirm the transcription factors identified

from network based modeling analysis of the microarray data in Calu3 2B4 cells (step #6 above), similar modeling approaches were performed to identify transcription factors in microarray data from icSARS-CoV and icSARS-CoV ΔORF6 infected primary human airway epithelial cell cultures (HAE; step #13 above) as well as from proteomic data from icSARS-CoV and icSARS-CoV ΔORF6 infected Calu3 2B4 (step #10 above). These additional experiments will further refine the differentially expressed genes or proteins networks and to independently validate the key core transcription factors as central regulators or hubs of the differentially expressed genes across primary and continuous lung epithelial cell lines. This will allow for comparison of significant regulatory transcription factor hubs that are important for blocking nuclear import during SARS-CoV infection and provides independent validation assessments between cell types, microarray and proteomic datasets.

Transcription factor analysis of the combined Calu3 2B4 proteomics and transcriptomics data resulted in an increase in enrichment scores, ranging from 1.5 to $2x10^{12}$-fold, for 22 out of 28 of the transcription factor hubs requiring karyopherin for nuclear importation, including 5 out of the 6 transcription factor hubs uniquely upregulated during icSARS-CoV ΔORF6 infection (**Table 2**). This suggests that many protein nodes in the proteomic dataset share common upstream transcription factor regulators with gene nodes in the transcriptional data, including those that require karyopherin for nuclear import.

For the HAE data, two of the enriched transcription factor hubs in HAE cultures, CREB1 and VDR, were also identified as important in the targeted Calu-3 transcriptomics analysis of the C4 cluster (**Table 1, Figures 3 and 4**) while the other 4 transcription

factors overlapped with the hubs identified from the global Calu-3 gene microarray data set (**Table 2**), independently validating the overlap of karyopherin-mediated transcriptional hubs between HAE and Calu3 2B4 cells.

15. Comparison of Creb1 and VDR transcription factor networks between Calu-3 and HAE cells (Figure 5B, Suppl. Figure S4). Karyopherin-mediated VDR and CREB1 transcription factor networks were significantly ($p<0.05$) enriched for both Calu-3 and HAE datasets. Nodes from HAE analysis and Calu-3 analysis include those unique to each cell type and some that are common between both cell types.

16. Validation of transcriptional targets between Calu-3 and HAE genes (Figure 5C) and Calu-3 genes and proteins (Figure 6). Comparison of gene expression in Calu-3 cells and HAE cells (post infection for target nodes downstream VDR and CREB1 are graphed in **Figure 5C**. The mRNA expression levels (from the microarray data) for individual genes that were differentially expressed in both HAE and Calu3 2B4 datasets include B cell translocation gene (Btg2), forkhead box 03a (Foxo3a), hypermethylated in cancer 2 (Hic2), human p-thromboglobulin gene (Ptg), thiamine transporter gene (Scl19a2), glucose transporter gene (Scl2a6), transforming growth factor beta 3 (TGFb3), and POK family transcription factor (Zbtb5). Transcription factor analysis in HAE cultures resulted in strong overlap with the transcription factor hubs identified in Calu3 2B4 cells and further supports the importance of karyopherin-mediated nuclear importation during SARS-CoV infection.

In addition, individual downstream protein nodes were augmented between the Calu3 2B4 transcriptomic and proteomic datasets, confirming transcriptional factor analysis.

Enhancer of mRNA-decapping enzyme 3 (EDC3) and Golgi adapter related complex of proteins, mu 1 subunit (AP3M1), two target gene nodes from the VDR and CREB1 networks, were significantly upregulated in icSARS-CoV ΔORF6 compared to icSARS-CoV as measured at both the transcript (graphed data is from microarray analysis) and protein (graphed data is from the proteomics data) level (**Figure 6**). Individual RNA expression values derived from microarray analysis demonstrate increased RNA expression trends in icSARS-CoV ΔORF6 versus icSARS-CoV in both EDC3 and AP3M1 **(Figure 6 A and C**). In addition, protein abundance increases at late times during icSARS-CoV ΔORF6 infection; in contrast, icSARS-CoV infection protein levels fail to rise above mock values (**Figure 6B and D**). When directly compared, both gene and protein analysis demonstrate augmentation of these gene nodes in icSARS-CoV ΔORF6 infection at late times when compared to icSARS-CoV. Together, these data confirm increased expression of targeted gene nodes downstream of identified transcriptional factors in icSARS-CoV ΔORF6 infection that are absent in icSARS-CoV infection.

1. **Cilloniz, C., M. J. Pantin-Jackwood, C. Ni, A. G. Goodman, X. Peng, S. C. Proll, V. S. Carter, E. R. Rosenzweig, K. J. Szretter, J. M. Katz, M. J. Korth, D. E. Swayne, T. M. Tumpey, and M. G. Katze.** 2010. Lethal dissemination of H5N1 influenza virus is associated with dysregulation of inflammation and lipoxin signaling in a mouse model of infection. J Virol **84:**7613-24.

2. **Dennis, G., Jr., B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane, and R. A. Lempicki.** 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol **4:**P3.

3. **Eng, J., A. L. McCormack, and J. R. Yates, 3rd.** 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom **5:**976-989.

4. **Frieman, M., B. Yount, M. Heise, S. A. Kopecky-Bromberg, P. Palese, and R. S. Baric.** 2007. Severe acute respiratory syndrome coronavirus ORF6 antagonizes STAT1 function by sequestering nuclear import factors on the rough endoplasmic reticulum/Golgi membrane. J Virol **81:**9812-24.

5. **Huang da, W., B. T. Sherman, and R. A. Lempicki.** 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc **4:**44-57.

6. **Jaitly, N., A. Mayampurath, K. Littlefield, J. N. Adkins, G. A. Anderson, and R. D. Smith.** 2009. Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. BMC Bioinformatics **10:**87.

7. **Kiebel, G. R., K. J. Auberry, N. Jaitly, D. A. Clark, M. E. Monroe, E. S. Peterson, N. Tolic, G. A. Anderson, and R. D. Smith.** 2006. PRISM: a data management system for high-throughput proteomics. Proteomics **6:**1783-90.

8. **Kopecky-Bromberg, S. A., L. Martinez-Sobrido, M. Frieman, R. A. Baric, and P. Palese.** 2007. Severe acute respiratory syndrome coronavirus open reading

frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists. J Virol **81:**548-57.

9. **Li, C., A. Bankhead, 3rd, A. J. Eisfeld, Y. Hatta, S. Jeng, J. H. Chang, L. D. Aicher, S. Proll, A. L. Ellis, G. L. Law, K. M. Waters, G. Neumann, M. G. Katze, S. McWeeney, and Y. Kawaoka.** 2011. Host regulatory network response to infection with highly pathogenic H5N1 avian influenza virus. J Virol **85:**10955-67.

10. **Livesay, E. A., K. Tang, B. K. Taylor, M. A. Buschbach, D. F. Hopkins, B. L. LaMarche, R. Zhao, Y. Shen, D. J. Orton, R. J. Moore, R. T. Kelly, H. R. Udseth, and R. D. Smith.** 2008. Fully automated four-column capillary LC-MS system for maximizing throughput in proteomic analyses. Anal Chem **80:**294-302.

11. **Matzke, M. M., K. M. Waters, T. O. Metz, J. M. Jacobs, A. C. Sims, R. S. Baric, J. G. Pounds, and B. J. Webb-Robertson.** 2011. Improved quality control processing of peptide-centric LC-MS proteomics data. Bioinformatics **27:**2866-72.

12. **Metz, T. O., J. M. Jacobs, M. A. Gritsenko, G. Fontes, W. J. Qian, D. G. Camp, 2nd, V. Poitout, and R. D. Smith.** 2006. Characterization of the human pancreatic islet proteome by two-dimensional LC/MS/MS. J Proteome Res **5:**3345-54.

13. **Monroe, M. E., N. Tolic, N. Jaitly, J. L. Shaw, J. N. Adkins, and R. D. Smith.** 2007. VIPER: an advanced software package to support high-throughput LC-MS peptide identification. Bioinformatics **23:**2021-3.

14. **Nikolsky, Y., E. Kirilov, R. Zuez, E. Rakhmatulin, and T. Nikolskaya.** 2009. Functional anlaysis of OMICs data and small molecule compounds in an integrated “knowledge-baseed” platform. *In* Y. N. a. J. Bryant (ed.), Protein Networks and Pathway Analysis, vol. 563. Humana Press, Totowa, New Jersey.

15. **Petritis, K., L. J. Kangas, P. L. Ferguson, G. A. Anderson, L. Pasa-Tolic, M. S. Lipton, K. J. Auberry, E. F. Strittmatter, Y. Shen, R. Zhao, and R. D. Smith.** 2003. Use of artificial neural networks for the accurate prediction of peptide liquid chromatography elution times in proteome analyses. Anal Chem **75:**1039-48.

16. **Petyuk, V. A., W. J. Qian, C. Hinault, M. A. Gritsenko, M. Singhal, M. E. Monroe, D. G. Camp, 2nd, R. N. Kulkarni, and R. D. Smith.** 2008. Characterization of the mouse pancreatic islet proteome and comparative analysis with other mouse tissues. J Proteome Res **7:**3114-26.

17. **Polpitiya, A. D., W. J. Qian, N. Jaitly, V. A. Petyuk, J. N. Adkins, D. G. Camp, 2nd, G. A. Anderson, and R. D. Smith.** 2008. DAnTE: a statistical tool for quantitative analysis of -omics data. Bioinformatics **24:**1556-8.

18. **Qian, W. J., T. Liu, M. E. Monroe, E. F. Strittmatter, J. M. Jacobs, L. J. Kangas, K. Petritis, D. G. Camp, 2nd, and R. D. Smith.** 2005. Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome. J Proteome Res **4:**53-62.

19. **Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker.** 2003. Cytoscape: a software

environment for integrated models of biomolecular interaction networks. Genome Res **13:**2498-504.

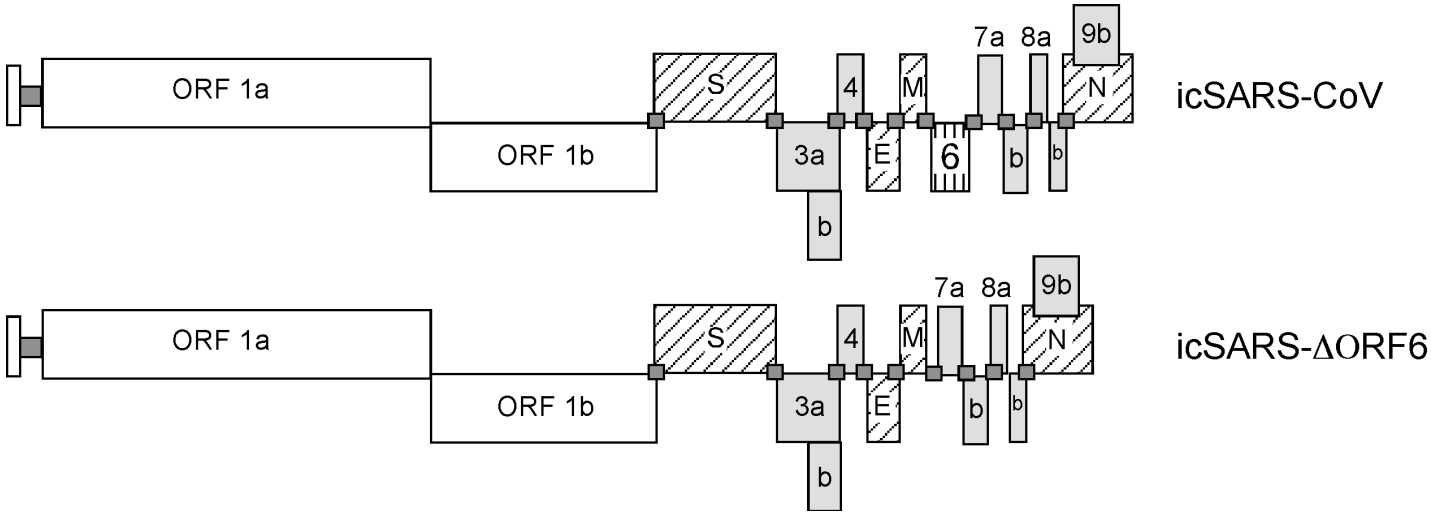
20. **Webb-Robertson, B. J., M. M. Matzke, J. M. Jacobs, J. G. Pounds, and K. M. Waters.** 2011. A statistical selection strategy for normalization procedures in LC-MS proteomics experiments through dataset-dependent ranking of normalization scaling factors. Proteomics **11:**4736-41.

21. **Webb-Robertson, B. J., L. A. McCue, K. M. Waters, M. M. Matzke, J. M. Jacobs, T. O. Metz, S. M. Varnum, and J. G. Pounds.** 2010. Combined statistical analyses of peptide intensities and peptide occurrences improves identification of significant peptides from MS-based proteomics data. J Proteome Res **9:**5748-56.

22. **Yount, B., R. S. Roberts, A. C. Sims, D. Deming, M. B. Frieman, J. Sparks, M. R. Denison, N. Davis, and R. S. Baric.** 2005. Severe acute respiratory syndrome coronavirus group-specific open reading frames encode nonessential functions for replication in cell cultures and mice. J Virol **79:**14909-22.

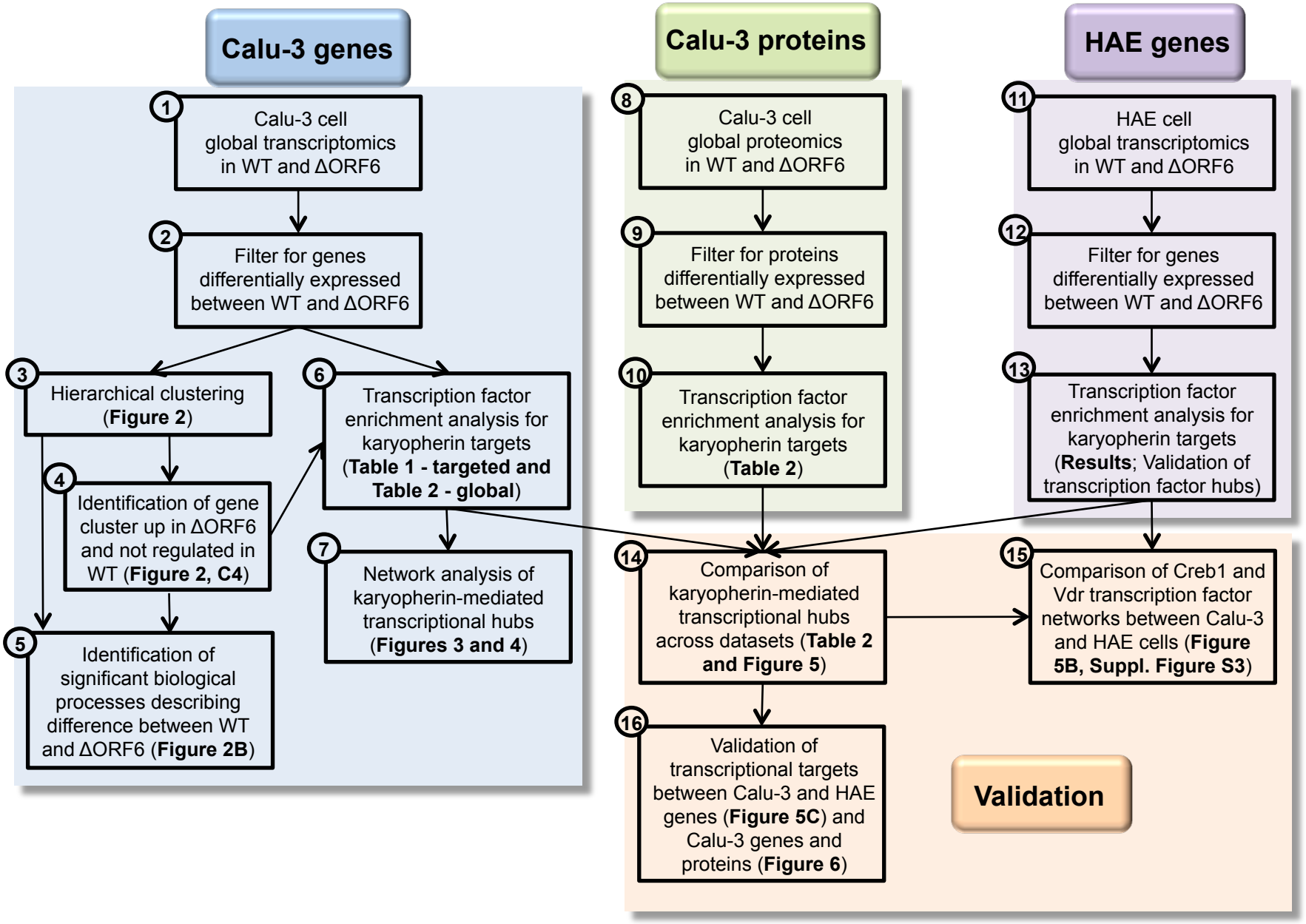
23. **Zhao, J., A. Falcon, H. Zhou, J. Netland, L. Enjuanes, P. Perez Brena, and S. Perlman.** 2009. Severe acute respiratory syndrome coronavirus protein 6 is required for optimal replication. J Virol **83:**2368-73.

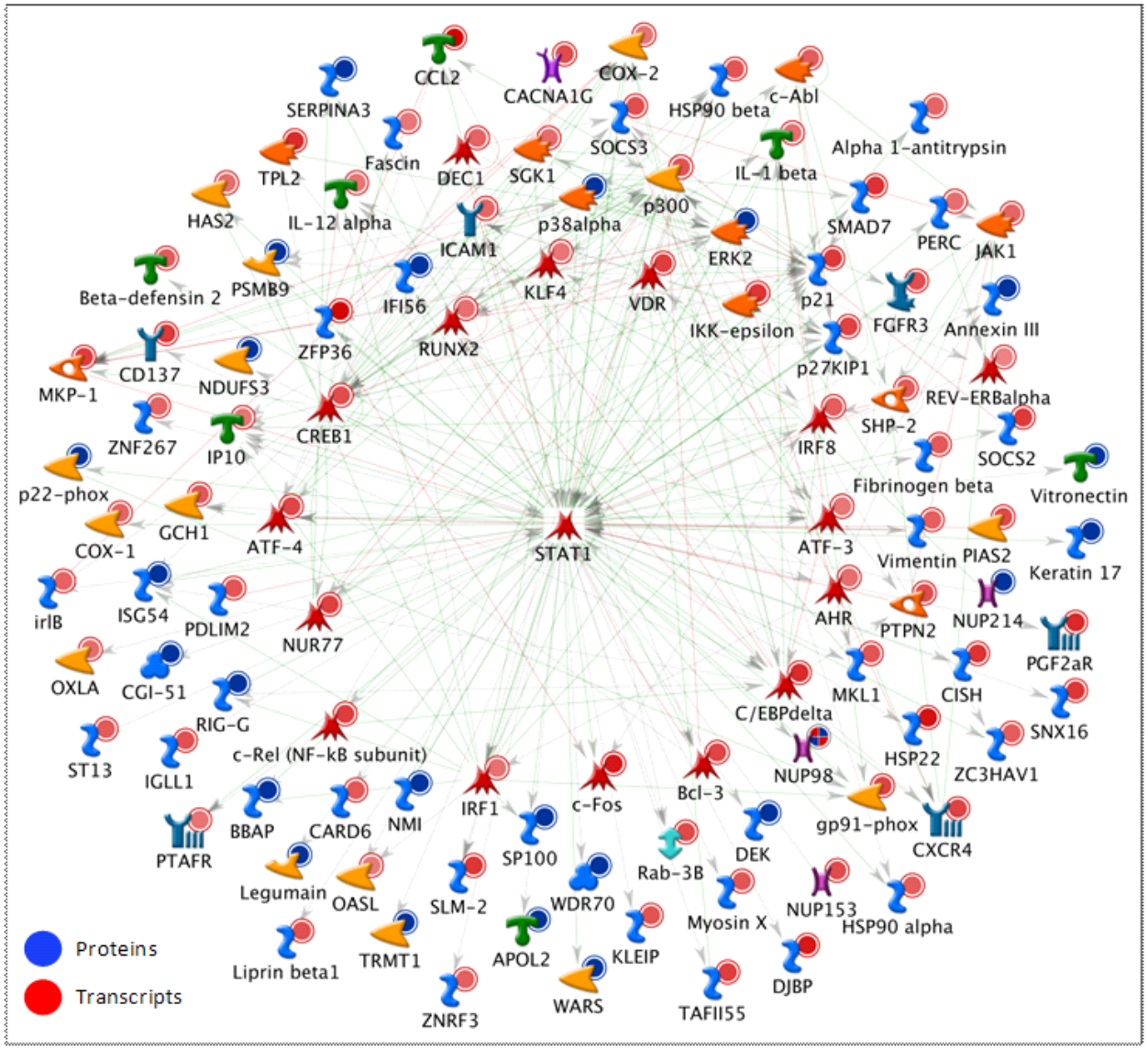
24. **Zimmer, J. S., M. E. Monroe, W. J. Qian, and R. D. Smith.** 2006. Advances in proteomics data analysis and display using an accurate mass and time tag approach. Mass Spectrom Rev **25:**450-82.



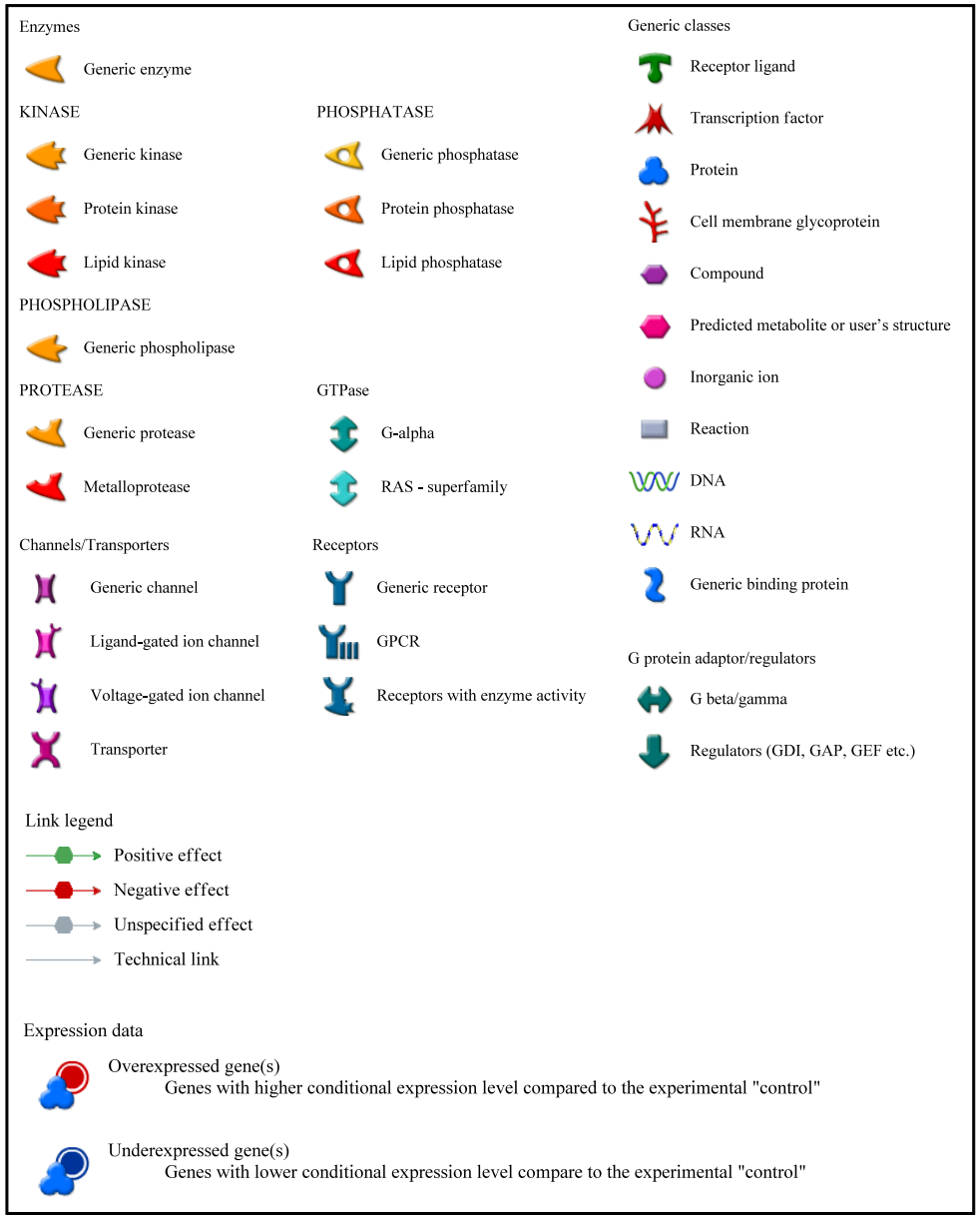
Supplemental Figure S1. icSARS-CoV ΔORF6 genome construction Shown are schematics of the wild type icSARS-CoV (top) and icSARS-CoV ΔORF6 (bottom) infectious clone constructs used in this study. Rectangles are used to represent the open reading frames (ORF) of the SARS-CoV genome, which can be divided into replicase (white), structural (striped), and accessory (shaded and spotted) ORFs. ORF6 the focus of this study is indicated by the spotted rectangle in the top schematic and has been deleted in the bottom construct (nucleotides 27,074-27265 were removed). Squares represent the transcriptional regulatory sequences that preceed the ORFs. White rectangle- replicase ORFs, striped rectangle- structural gene ORFs, shaded rectangle- accessory ORFs

Supplemental Figure S2: Summary of modeling workflow. Steps are described in detail in Supplemental Section S2.

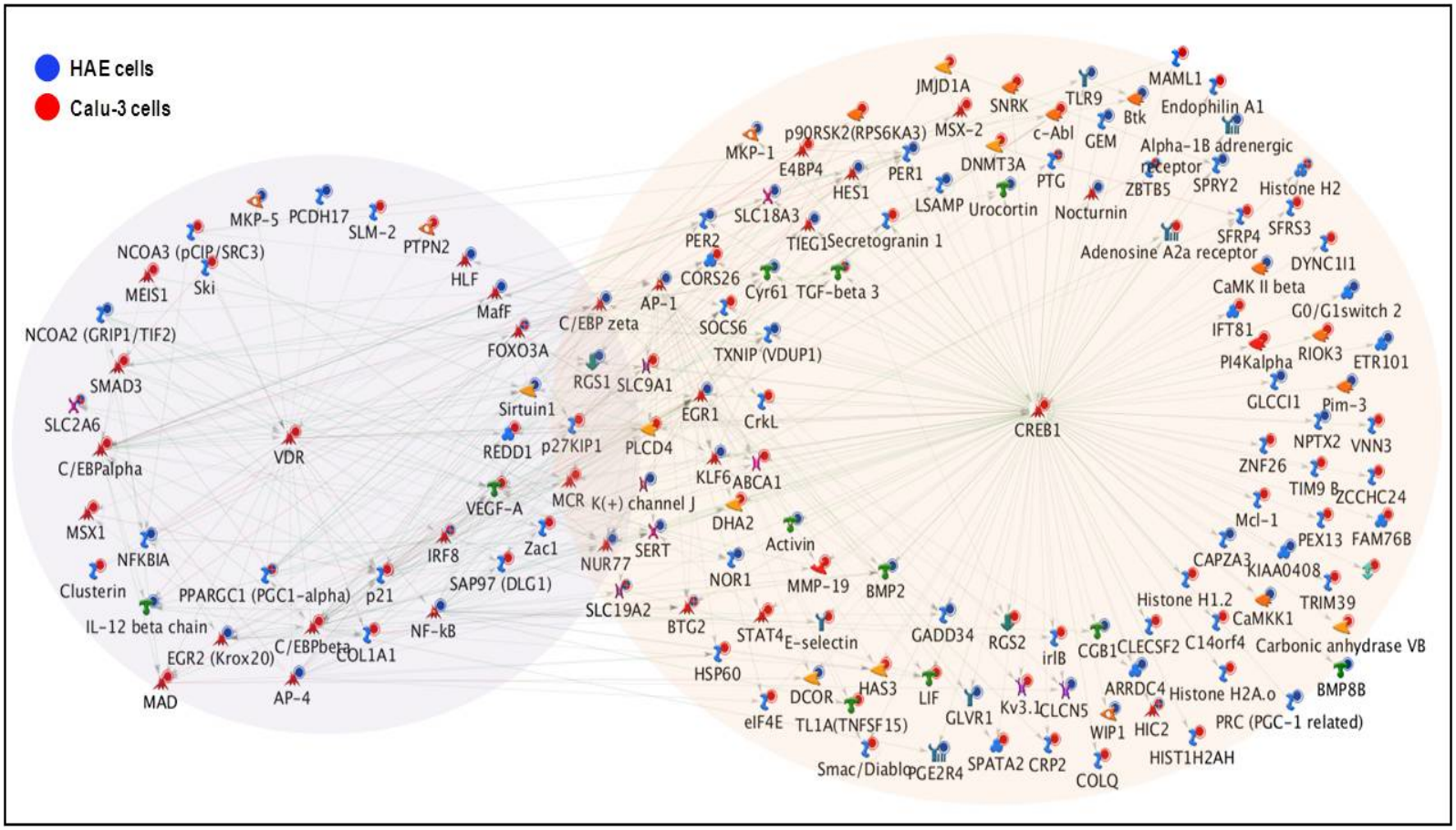




Supplemental Figure S3A. STAT1 network. A. Transcription factor network for STAT1 showing all connected target nodes in Calu-3 dataset that are upregulated in icSARS-CoV ΔORF6 compared to icSARS-CoV. Nodes from proteomic analysis are blue and transcriptomic analysis are red. **B.** MetaCore network legend.



Supplemental Figure S3B. STAT1 network. A. Transcription factor network for STAT1 showing all connected target nodes in Calu-3 dataset that are upregulated in icSARS-CoV ΔORF6 compared to icSARS-CoV. Nodes from proteomic analysis are blue and transcriptomic analysis are red. **B.** MetaCore network legend.



Supplemental Figure S4. HAE and Calu-3 cells were infected with wild type and SARS-CoV ΔORF6 (MOI) and were harvested at 24-72 hr (HAE) or 0-72 hr (Calu-3) post infection for microarray analysis. Karyopherin-mediated VDR and CREB1 transcription factor networks were significantly ($p<0.05$) enriched for both Calu-3 and HAE datasets. Nodes from HAE analysis are blue and Calu-3 analysis are red.